WILEY | Hindawi

Retraction

Retracted: Vehicle Motion State Prediction Method Integrating Point Cloud Time Series Multiview Features and Multitarget Interactive Information

Journal of Advanced Transportation

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] R. Zhang, Y. Guo, Y. Long, Y. Zhou, and C. Jiang, "Vehicle Motion State Prediction Method Integrating Point Cloud Time Series Multiview Features and Multitarget Interactive Information," *Journal of Advanced Transportation*, vol. 2022, Article ID 4736623, 21 pages, 2022.

WILEY | Hindawi

Research Article

Vehicle Motion State Prediction Method Integrating Point Cloud Time Series Multiview Features and Multitarget Interactive Information

Ruibin Zhang [1,2] **Yingshi Guo** [1] **Yunze Long**,[2] **Yang Zhou** [1] **and Chunyan Jiang**[2]

[1]*School of Automobile, Chang'an University, Xi'an 710064, China*
[2]*School of Automobile Engineering, Guilin University of Aerospace Technology, Guilin 541004, China*

Correspondence should be addressed to Ruibin Zhang; 2016022016@chd.edu.cn

A vehicle motion state prediction algorithm integrating point cloud timing multiview features and multitarget interaction information is proposed in this work to effectively predict the motion states of traffic participants around intelligent vehicles in complex scenes. The algorithm analyzes the characteristics of object motion that are affected by the surrounding environment and the interaction of nearby objects and is based on the complex traffic environment perception dual multiline light detection and ranging (LiDAR) technology. The time sequence aerial view map and time sequence front view depth map are obtained using real-time point cloud information perceived by the LiDAR. Time sequence high-level abstract combination features in the multiview scene are then extracted by an improved VGG19 network model and are fused with the potential spatiotemporal interaction of the multitarget operation state data extraction features detected by the laser radar by using a one-dimensional convolution neural network. A temporal feature vector is constructed as the input data of the bidirectional long-term and short-term memory (BiLSTM) network, and the desired input-output mapping relationship is trained to predict the motion state of traffic participants. According to the test results, the proposed BiLSTM model based on point cloud multiview and vehicle interaction information is better than other methods in predicting the state of target vehicles. The results can provide support for the research to evaluate the risk of intelligent vehicle operation environment.

1. Introduction

The effective motion state prediction of surrounding traffic participants by intelligent vehicles in complex scenes is an important link to realize driverless technology. In this work, microlevel data, such as the driving track, velocity, and acceleration of the moving target, are obtained using sensor equipment, including light detection and ranging (LiDAR) and cameras. Employing data from research fields that include state estimation [1], intention recognition [2], trajectory prediction [3], intelligent driving [4], driving behavior analysis [5], and safety risk detection [6] can assist intelligent transportation systems in improving traffic safety and reduce accidents.

Dynamic object state prediction can predict an object's future state by using its perceived historical state information. At present, mainstream methods mainly include traditional, machine learning, and depth learning methods. Traditional methods are largely based on the assumption of object kinematics, where prediction results are calculated by establishing kinematic or dynamic models to estimate the propagation of object motion state with time. Ammoun and Nashashibi [7] used a linear Kalman filter to estimate and propagate future states and predict trajectory by constructing a vehicle dynamics model. The state estimation included position, velocity, and acceleration [8, 9]. Kim and Yi [10] defined the expected yaw rate required for lane change and curve and added this to the extended Kalman filter [11]. In order to further improve prediction accuracy, Schreier et al.

[12] proposed the probabilistic trajectory prediction method of Monte Carlo simulation. However, these methods fail to accurately capture the complex and changeable motion characteristics of dynamic objects and can only obtain accurate results in a very short time. As they do not match the real trajectory in predictions of more than 1 second, such methods may not be effective in practical applications. Trajectory prediction methods based on machine learning mainly include Gaussian process, hidden Markov model, or Bayesian network [13–15]. Tran et al. [16–18] learned the model parameters of vehicle trajectory through a Gaussian process, and Patterson et al. [19, 20] used a mixed Gaussian model to learn the vehicle trajectory generation model. Streubel et al. [21, 22] independently predicted the discrete action of each object using a hidden Markov model. However, in real scenes, the assumption of total independence is typically unrealistic. Gindele et al. [23] formulated a more complex model based on vehicle interaction by using a dynamic Bayesian network to predict vehicle trajectory, but the network was more expensive to calculate. These methods become complicated when processing high-dimensional data and require manually designed input features to capture context information. This limits the flexibility of learning algorithms, resulting in poor performance. In addition, such methods can only predict the behavior of specific entities. With the success of deep learning in various computer vision and robotics fields, many researchers have begun to introduce deep learning into trajectory prediction tasks [24]. Most trajectory prediction methods based on deep learning use the recurrent neural network (RNN) and its variant structure long short-term memory (LSTM) or gated recursive unit (GRU) to model the behavior of objects. For example, Kim et al. [25] built an aerial map grid using the LSTM model to establish an LSTM structure for each vehicle and employed past trajectory data to predict the position of the vehicle in the grid in the next 2 s. Deo et al. [26] used the LSTM encoding decoding structure to first identify the vehicle maneuvers (left lane change, straight lane change, and right lane change) and then carried out multimodal prediction based on different maneuvers. As these methods regard each vehicle as a separate object for independent prediction, they are only suitable for high-velocity scenes with simple motion modes [27–30]. In complex urban scenes, the future movement of targets and cars is affected by the movement of other objects and the spatial environment. To improve the accuracy of trajectory prediction by addressing complexity, researchers have begun to model the social interaction between multiple objects and the constraints of scene context based on object trajectory prediction [31–35]. Alahi et al. [36] proposed the social LSTM model, which captured the social interaction of the target by running a maximum pool of the state vector of the nearby target within a predefined distance range. However, this interaction did not model the social interaction of the distant target. Vemula [37] proposed a social attention model, which did not require the restriction of the local neighborhood hypothesis, and predicted target trajectories based on the social interactions of all targets in

space through attention mechanism and spatiotemporal mapping [38]. Sadeghian et al. [39] proposed an attention recurrent neural network, which used the past motion trajectory of the target and the top view image of the navigation scene as inputs, and obtained more accurate prediction results by learning the influence of the spatial environment on the target trajectory. Haddad et al. [40] used a spatiotemporal graph model to encode the motion influence of static objects in the scene on targets and the interaction between targets. Sadeghian et al. [41] proposed the SoPhie model, which used an attention mechanism to help the model extract the most prominent part in the image related to the path and the interaction information between different targets to predict the motion trajectory of the target. Given the current research progress, some achievements have been made in moving target state prediction for urban scenes, most studies only consider the interaction between objects, and the context information of the running scene is obtained from a single perspective of the sensor. There is limited research on the method of moving target state prediction integrating environmental view features and multitarget interactive information. This paper presents a vehicle motion state prediction method that integrates the multiview characteristics of point cloud timing and multiobjective interactive information by analyzing the characteristics that object motion in urban scenes is affected by the interaction of other surrounding objects and the surrounding environment. Our contributions are mainly given by

(1) Vehicle mounted dual multiline LiDAR data acquisition system. In order to effectively obtain the interaction information between the surrounding environment of intelligent vehicles and traffic participants in complex traffic scenes, an on-board dual multiline LiDAR environment sensing technology is proposed. The point cloud data collected by the LiDAR on the upper side of the on-board mounting bracket are used to obtain the global environmental map information in real time. The point cloud collected by the LiDAR on the lower side is used to detect the operation status data of traffic participants in real time.

(2) Point cloud multiview environment feature extraction network. In order to effectively extract the characteristics of the surrounding environment of intelligent vehicles in complex traffic scenes, two improved VGG19 network models are used to extract the multiview features of point cloud. One improved VGG19 network branch is used to extract the features of point cloud depth map, and the other improved VGG 19 network branch is used to extract the features of point cloud aerial view map.

(3) Multiobjective interactive information extraction network. In order to effectively extract the interaction relationship between intelligent vehicles and surrounding target vehicles in complex traffic scenes, the multitarget historical motion state data obtained

by lidar are input into 1dcnn network, and the network outputs the potential space-time interaction relationship between vehicles.

(4) A vehicle motion state prediction model integrating multiview features of point cloud and multitarget interactive information. The output features of the point cloud multiview feature extraction network and the multitarget interactive information extraction network are fused, and the BiLSTM model and the BiLSTM network structure are input. The long-term dependence between sequences is captured through the forward-backward network, and then, the motion state of the target vehicle is predicted.

2. Background

In order to effectively collect the interaction information between the surrounding environment of intelligent vehicles and traffic participants in complex traffic scenes, an environment sensing technology based on dual multiline LiDAR is proposed in this paper. The installation position of the dual multiline LiDAR system is shown in Figure 1, where the LiDAR adopts the vertical installation mode. The LiDAR located on the upper side of the vehicle is used to obtain the global environmental map information in real time and generate the time sequence top view aerial view map and time sequence front view depth map. The LiDAR located at the lower side of the vehicle is used to detect the operation status data of traffic participants in real time and obtain the complex interaction information between traffic participants.

The mathematical model of the installation position of the dual multiline LiDAR system is as follows:

$$\begin{cases} d_{i_{\uparrow}} = h_{\uparrow} \times \tan\left(\alpha_{\uparrow} + i_{\uparrow} \times \Delta\varnothing_{\uparrow}\right), \\ d_{i_{\downarrow}} = h_{\downarrow} \times \tan\left(\alpha_{\downarrow} + i_{\downarrow} \times \Delta\varnothing_{\downarrow}\right), \\ d_{\uparrow} = h_{\uparrow} \times \tan\left(\alpha_{\uparrow}\right), \\ d_{\downarrow} = h_{\downarrow} \times \tan\left(\alpha_{\downarrow}\right), \end{cases} \tag{1}$$

where ↑ represents the radar on the upper side in the dual LiDAR system, ↓ represents the radar on the lower side in the dual LiDAR system, and h is the height of the center point of the LiDAR relative to the ground. α is the included angle between the horizontal line of the LiDAR center and the lowest scanning line, $\Delta\varnothing$ is the vertical angular resolution of the LiDAR, d is the distance between the ground intersection of the lowest scanning line and the ground projection point of the LiDAR center, and d_i is the projection distance of the scanning line above the i-th horizontal plane of the LiDAR on the ground.

2.1. Point Cloud Multiview Generation.
Detection accuracy is usually low in complex environments because the images collected by the camera lack accurate depth information. Although the LiDAR point cloud has accurate depth information, in the method based on LiDAR point cloud, the point cloud is sparse, can only achieve high-precision three-dimensional

frame positioning, has poor detection effect on small objects, and is prone to missed detection or false detection. Comparatively, the aerial view has three advantages. First, the objects in the aerial view occupy different spaces to avoid the occlusion problem. Second, when the object is projected onto the aerial view, the physical size will be retained, so the change of size will be very small. Third, the location of the aerial view is critical for obtaining an accurate 3D boundary. The forward-looking depth map can obtain the depth information of the area in front of the moving object. Therefore, based on the multiview method of point cloud, this paper transforms the original point cloud of LiDAR into aerial view, forward-looking depth map, and other image forms for processing to provide necessary high-precision environmental perception information inputs for subsequent moving target state prediction.

The multiview generation process of point cloud is used to convert the original LiDAR point cloud into a top aerial view and front depth map, as shown in Figure 2.

The aerial view and depth map obtained from the original point cloud of 3D LiDAR need to go through the process of 2D projection of the 3D image, which should be transformed according to the internal parameters of the corresponding camera. The conversion relationship from the LiDAR coordinate system to the image coordinate system is as follows:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \dfrac{1}{dx} & 0 & u_0 \\ 0 & \dfrac{1}{dx} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R_{L-C} & T_{L-C} \\ 0^T & 1 \end{bmatrix} \cdot \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix}, \tag{2}$$

where (u, v) is the pixel coordinate, (X_C, Y_C, Z_C) is the camera coordinate, (X_L, Y_L, Z_L) is the LiDAR coordinate, R_{L-C} represents the rotation matrix from the LiDAR coordinate system to the camera coordinate system, and T_{L-C} represents the three-dimensional translation vector from the LiDAR coordinate system to the world coordinate system. u_0, v_0 is the internal parameter of the camera, f is the focal length of the camera, and Z_C is the depth value corresponding to the current image coordinate $(u, v, 1)^T$.

2.2. Multiobjective Information Interaction Network for Complex Traffic Scenes.
Target vehicle trajectory prediction in complex traffic scenes works to estimate the future state according to its own running state, combined with the spatiotemporal interaction relationship of the surrounding environment. In order to effectively mine the interactive features in complex traffic data, the input data should include parameters of vehicle size, operation state, and space-time relationship between vehicles.

The multitarget information interaction network in a complex traffic scene is shown in Figure 3. The red vehicle is an intelligent vehicle equipped with a LiDAR sensor, and its operation state model is as follows:
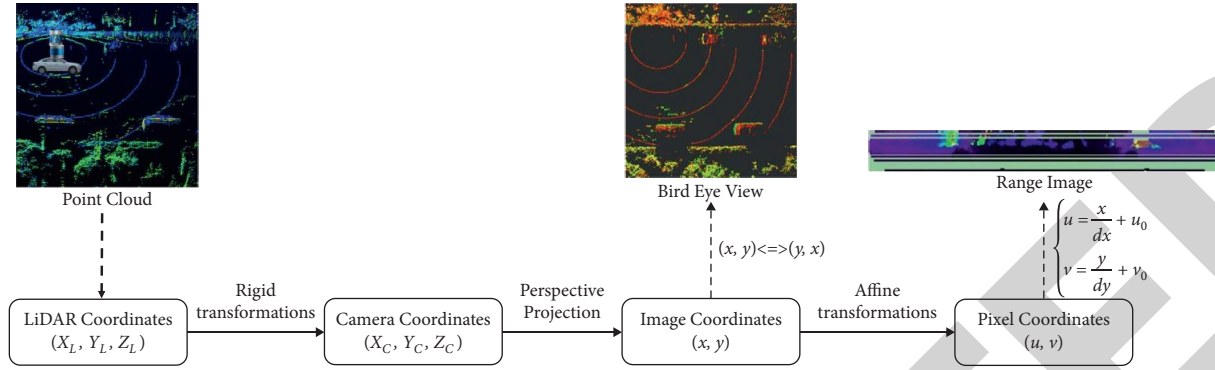
FIGURE 1: Installation position of the dual multiline LiDAR system.

$$O^{(t)} = \left[L, W, X^{(t)}, Y^{(t)}, V^t, \left(X^{(t-5)}, Y^{(t-5)} \right) \ldots \left(X^{(t-1)}, Y^{(t-1)} \right) \right], \tag{3}$$

where L is the length of the intelligent vehicle, W is the width of the intelligent vehicle, and $X^{(t)}, Y^{(t)}$ are the horizontal and vertical axis coordinates of the center point of the onboard LiDAR, respectively. $V^{(t)}$ is the instantaneous velocity of the intelligent vehicle, and $(X^{(t-5)}, Y^{(t-5)}) \ldots (X^{(t-1)}, Y^{(t-1)})$ are historical tracks.

The green vehicle O_1 in front of the adjacent right is perceived by the onboard LiDAR of the intelligent vehicle. The blue vehicle O_2 is at the rear right, and the yellow rear vehicle O_3 is a surrounding vehicle whose state behavior needs to be predicted. The green vehicle O_1 in the front right position is an example of the target, and its operation state model is as follows:

$$O^{(t)} = \left[L_1, W_1, X_1^{(t)}, Y_1^{(t)}, a_1^{(t)}, V_1^{(t)}, \psi_1^{(t)}, \omega_1^{(t)}, \left(X_1^{(t-5)}, Y_1^{(t-5)} \right) \ldots \left(X^{(t-1)}, Y^{(t-1)} \right) \right], \tag{4}$$

where L_1 is the target vehicle length, W_1 is the target vehicle width, $X_1^{(t)}, Y_1^{(t)}$ is the horizontal and vertical axis coordinate value of the center point of the target vehicle, $a_1^{(t)}$ is the instantaneous acceleration of the target vehicle, $V_1^{(t)}$ is the instantaneous velocity of the target vehicle, $\psi_1^{(t)}$ is the instantaneous yaw angle, and $(X_1^{(t-5)}, Y_1^{(t-5)}) \cdots (X_1^{(t-1)}, Y_1^{(t-1)})$ are historical tracks.

The red smart car and the green vehicle in the front right position are referred to as O_1, the left rear blue vehicle is O_2, and the yellow rear vehicle is O_3. The multiobjective information interaction network model of complex traffic scene is as follows:

$$O^{(t)} = \left[O_1, O_1^{(t)}, O_2^{(t)}, O_3^{(t)} \cdots \left(X_1^{(t)}, Y_1^{(t)} \right), \left(X_2^{(t)}, Y_2^{(t)} \right), \left(X_3^{(t)}, Y_3^{(t)} \right) \right] \cdots, \tag{5}$$

where $X_1^{(t)}, Y_1^{(t)}$ refer to the horizontal and vertical axis coordinate values of the nearest point cloud coordinate point of the green vehicle in front of the right relative to the center point of the onboard LiDAR; $X_2^{(t)}, Y_2^{(t)}$ refer to the horizontal and vertical axis coordinates of the nearest point cloud

coordinates of the left rear blue vehicle relative to the center point of the onboard LiDAR; and $X_3^{(t)}, Y_3^{(t)}$ refer to the horizontal and vertical axis coordinates of the nearest point cloud coordinates of the yellow rear vehicle relative to the center point of the onboard LiDAR.

FIGURE 2: Point cloud multiview generation process.

3. Methodology

3.1. Multiview Feature Extraction Network.

To overcome data potential limitations [42], we propose a multiview feature fusion network. Two improved VGG19 network models are used to extract point cloud multiview features in this work. One branch extracts point cloud depth map features, and the other branch extracts point cloud aerial view features. The improved VGG19 network model adds several convolution layers on the basis of a shallow convolution neural network (CNN). Since adding a convolution layer is more conducive to image feature extraction than adding a full connection layer, the improved VGG19 network model can more easily overcome the lack of diversity and complexity of traffic scenes than a shallow convolution neural network and ultimately achieves a better spatiotemporal feature extraction effect. As shown in Figure 4, the input layer image size is 224×224 pixels, the number of channels is 3, and the VGG19 model has 16 convolution layers in total. There is a max-pooling layer behind the convolution layers 2, 4, 8, 12, and 16. The convolution kernel size in the convolution layer ranges from 224×224 reduced by half to 14×14. In this way, the use of progressively decreasing convolution kernel is equivalent to adding implicit regularization, which can improve the feature extraction ability of the network and increase its operation velocity. The improved VGG19 has four full connection layers, and the number of neurons is 4096, 4096, 1000, and 5, respectively. The last layer splices and outputs the high-level abstract combination features under the multiview scene extracted from the corresponding time-series point cloud depth map and point cloud aerial view.

3.2. Multiobjective Interactive Information Extraction Network.

Convolution neural network (CNN) is based on convolution calculations. Unlike manual design and feature extraction, CNN can automatically extract deep-level features. In CNN, the input data are first transformed by a series of chain convolution kernels with a nonlinear activation function, which is equivalent to the application of a series of chain multichannel nonlinear filters. The network learns the complex characteristics of data by stacking multiple convolution layers and using a nonlinear activation function.

The multiobjective interaction relationship corresponds to one-dimensional time series, so 1DCNN can be used to extract the potential interaction relationship. Specifically, the corresponding local information can be calculated by sliding the convolution kernel of a specific size through the local area of the input data. Convolution layer is the core of 1DCNN network, including convolution and excitation operation. Using convolution check, all image pixels on the input point cloud multiview go through it, and to add an offset coefficient b_i^l to the output image, the characteristic image is obtained by nonlinear transformation with excitation function. The mathematical expression of convolution layer operation is as follows:

$$Z^l = f\left(\sum_{i \in M} x_i^{l-1} k_i^l + b_i^l \right), \quad (6)$$

where l is the l-th convolution layer; Z^l is the l-th output layer L; b_i^l is the l-th offset of convolution kernel; x_i^{l-1} is the l-th input layer; f is the activation function; and k_i^l is the weight. In order to avoid the disappearance of gradient, alleviate the occurrence of over fitting, accelerate the convergence speed, and improve the accuracy, *ReLU* is selected as the activation function in this paper.

As shown in Figure 5, the multitarget interactive information extraction network inputs the LiDAR to obtain the multitarget historical motion state data of the time sequence frame of the complex traffic scene, outputs the potential space-time interaction diagram of the corresponding time sequence frame extracted through the 1DCNN network, and represents each dynamic target with a node. The nodes corresponding to any two targets in the same point cloud frame are connected with a solid line to represent the space edge, and the same target in adjacent frames is connected with a dotted line to represent the time edge.

4. Model Construction

4.1. Model Framework.

The overall network architecture of the vehicle motion state prediction model integrating point cloud multiview features and multiobjective interactive information proposed in this paper is shown in Figure 6. The network is mainly composed of a point cloud multiview
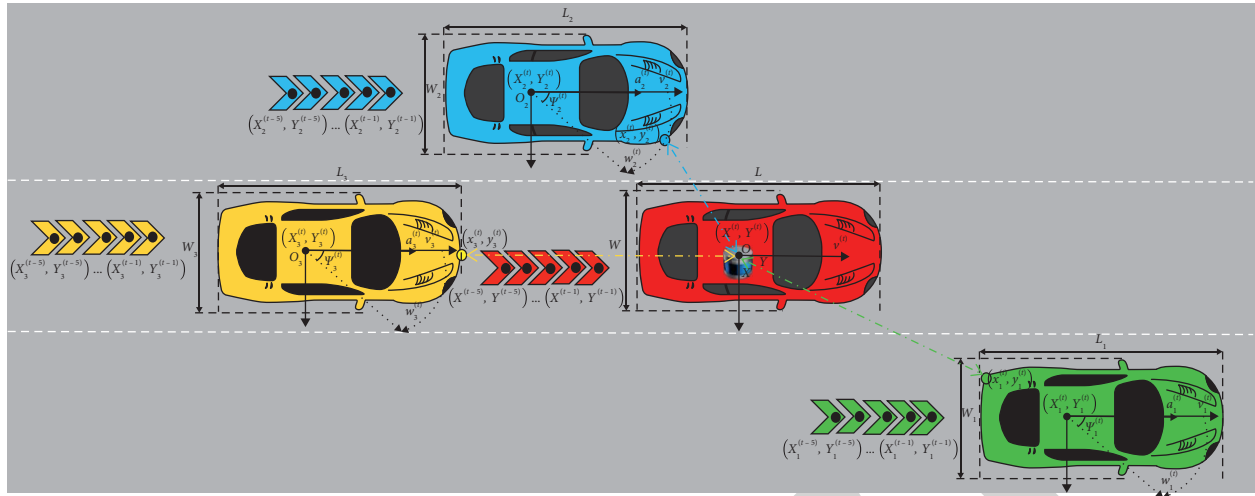
Figure 3: Multiobjective information interaction network in a complex traffic scene.
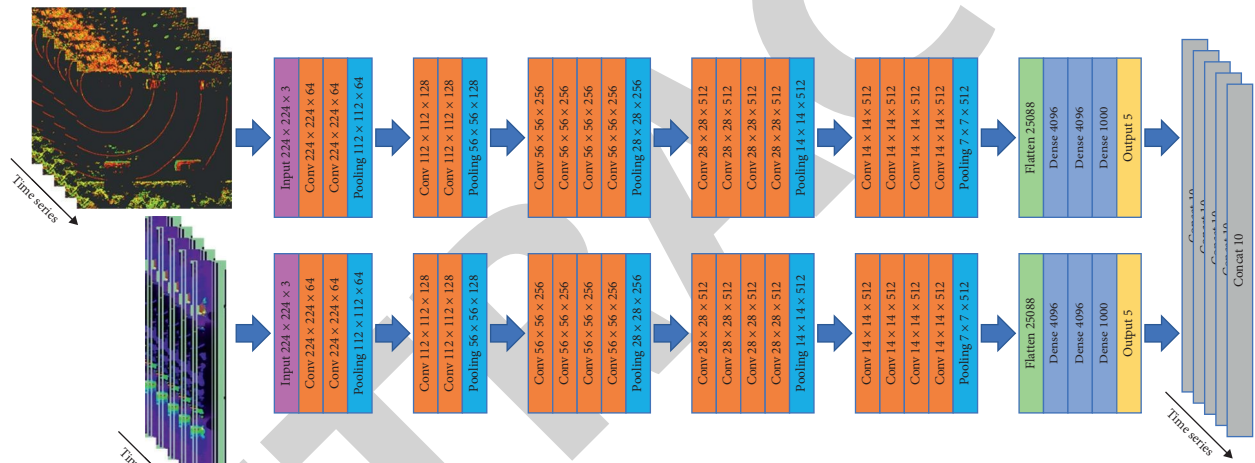

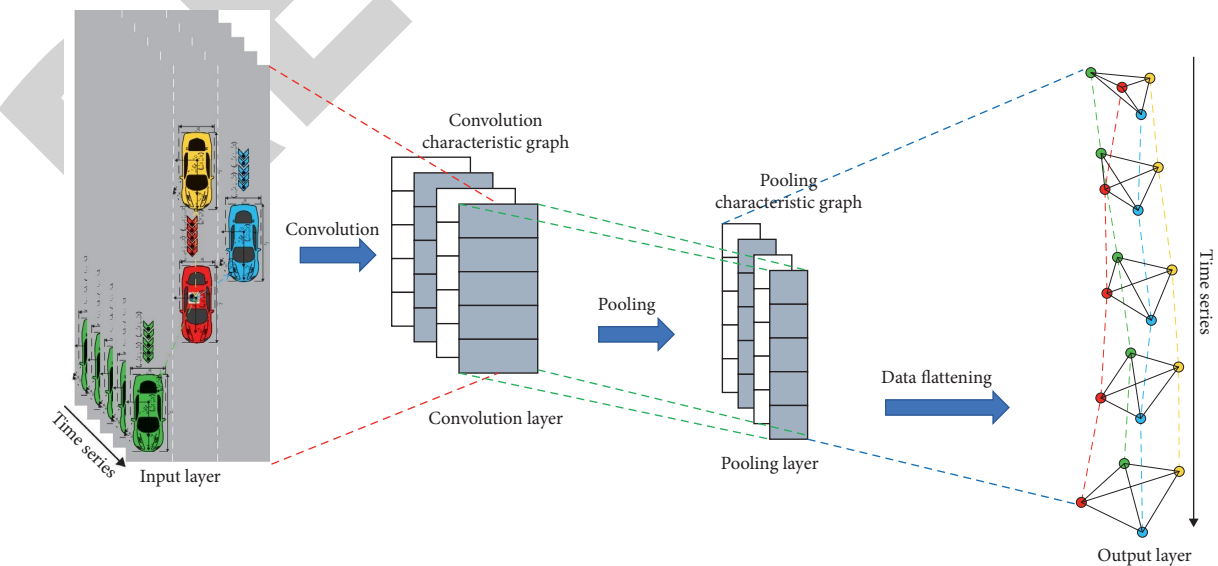Figure 4: Multiview feature extraction network.

Figure 5: Multiobjective interactive information extraction network.

feature extraction network, multiobjective interactive information extraction network, and a two-way long-term and short-term memory network. By inputting the historical motion state of multiple targets in the complex traffic scene obtained by sensors such as LiDAR and the corresponding time-series point cloud aerial view and point cloud forward-looking depth map, the network outputs the motion state of multiple targets in the future.

4.2. Model Implementation. Because LSTM network can learn and remember useful information and forget useless information through the training process to capture the dependence between the front and back long-distance time series traffic scene data, but it cannot encode the back and forward information. The BiLSTM [43] layer is composed of forward and backward LSTM layers. LSTM is an evolution of the traditional RNN network. Its internal memory unit can indicate the forgetting time of historical information and the update time of new information. LSTM can effectively improve the problems of gradient explosion and gradient disappearance caused by backpropagation errors in the network learning process. Moreover, LSTM can effectively learn sequence information by capturing long-distance dependencies. The structure of LSTM is shown in Figure 7.

The mathematical model of the LSTM structural unit is given by

$$
\begin{cases}
f_i^{(\tau)} = \sigma\left(b_i^f + \sum_j U_{i,j}^f x_j^{(\tau)} + \sum_j W_{i,j}^f h_j^{(\tau-1)}\right) \\
S_i^{(\tau)} = f_i^{(\tau)} S_i^{(\tau-1)} + g_i^{(\tau)}\sigma\left(b_i + \sum_j U_{i,j} x_j^{(\tau)} + \sum_j W_{i,j} h_j^{(\tau-1)}\right) \\
g_i^{(\tau)} = \sigma\left(b_i^g + \sum_j U_{i,j}^g x_j^{(\tau)} + \sum_j W_{i,j}^g h_j^{(\tau-1)}\right) \\
h_i^{(\tau)} = \tanh\left(S_i^{(\tau)} q_i^{(\tau)}\right) \\
q_i^{(\tau)} = \sigma\left(b_i^q + \sum_j U_{i,j}^q x_j^{(\tau)} + \sum_j W_{i,j}^q h_j^{(\tau-1)}\right)
\end{cases}
\tag{7}
$$

where $f_i^{(\tau)}$ is the output of the forget gate, $S_i^{(\tau)}$ is update the status in the cell, $g_i^{(\tau)}$ is the gate value entered for the current time, $h_i^{(\tau)}$ is the output of the current time, $h_j^{(\tau-1)}$ is the output of the previous time, $q_i^{(\tau)}$ is the output gate, and $x_j^{(\tau)}$ is the input vector for the current time. Moreover, b_i^f, $U_{i,j}^f$, and $W_{i,j}^f$ are, respectively, the input bias, input weight, and cycle weight of the forget gate; b_i^q, $U_{i,j}^q$, and $W_{i,j}^q$ are, respectively, the input bias, input weight, and cycle weight of the output gate; and b_i^g, $U_{i,j}^g$, and $W_{i,j}^g$ are, respectively, the input bias, input weight, and cycle weight of the input gate. Finally, tanh is a hyperbolic tangent function, σ is the sigmoid activation function, and i, j is the number of neurons, $i, j = 1, 2, \cdots, n$.

Step 1. The output features of the point cloud multiview feature extraction network and the multitarget interactive information extraction network are fused to obtain the feature vector $I^{(\tau)}$ at the current time;

Step 2. To effectively reduce the problem of large network prediction error caused by different orders of magnitude and feature extraction methods between the original data, the feature vector $I^{(\tau)}$ is normalized by the formula, that is, Z-score standardization, and the data are transformed into a dimensionless form.

$$
w_{out} = \frac{I^{(t)} - mean}{std},
\tag{8}
$$

where w_{out} is the dimensionless data, *mean* is the mean of the original data, and *std* is the standard deviation of the original data.

Step 3. The dimensionless data w_{out} are added into the bidirectional long-term and short-term memory (BiLSTM) network structure (shown in Figure 8) to capture the long-term dependency between sequences through the forward and backward network.

In (9), x_t is w_{out} eigenvectors in the output data, b_y is the offset of the hidden unit of the eigenvector in the BiLSTM network. $\overrightarrow{h_t}$ is an LSTM unit composed of feature sequences input from front to back, $\overleftarrow{h_t}$ is an LSTM unit composed of feature sequences input from back to front, y_t is the corresponding output result after the eigenvector passes through the BiLSTM network, $W_{\overrightarrow{h_y}}$ is the weight from the forward LSTM layer to the output layer, and $W_{\overleftarrow{h_y}}$ is the weight from the input layer to the backward LSTM layer [44].

$$
\begin{cases}
\overrightarrow{h_t} = LSTM\left(x_t, \overrightarrow{h_{t-1}}\right) \\
\overleftarrow{h_t} = LSTM\left(x_t, \overleftarrow{h_{t+1}}\right) \\
y_t = W_{\overrightarrow{h_y}}\overrightarrow{h_t} + W_{\overleftarrow{h_y}}\overleftarrow{h_t} + b_y
\end{cases}
\tag{9}
$$

Step 4. After effective fitting, the data y_t output the target state prediction value through two full connection layers in turn.

5. Experiments and Results

In order to verify the effectiveness of the vehicle motion state prediction method integrating point cloud timing multiview features and multitarget interactive information, an intelligent vehicle experimental platform is used for data collection, as illustrated in Figure 9. The experimental platform vehicle is a Shanghai Volkswagen Langyi 2013 1.6 L automatic comfort version, measuring 46 05 mm × 1765 mm × 1460 mm (length × wide × height). The platform includes RS-LiDAR-16 LiDAR, RS-LiDAR-32 LiDAR, a Gigabit Ethernet switch, algorithm processor, notebook computer, uninterrupted power supply, and other equipment. The 16 line LiDAR can scan the surrounding environment with a vertical field of view angle of −15°~15° and a horizontal angle of view of 360°, with a maximum ranging range of 150 m and an output of 32×10^4 points per second, with the scanning frequency set to
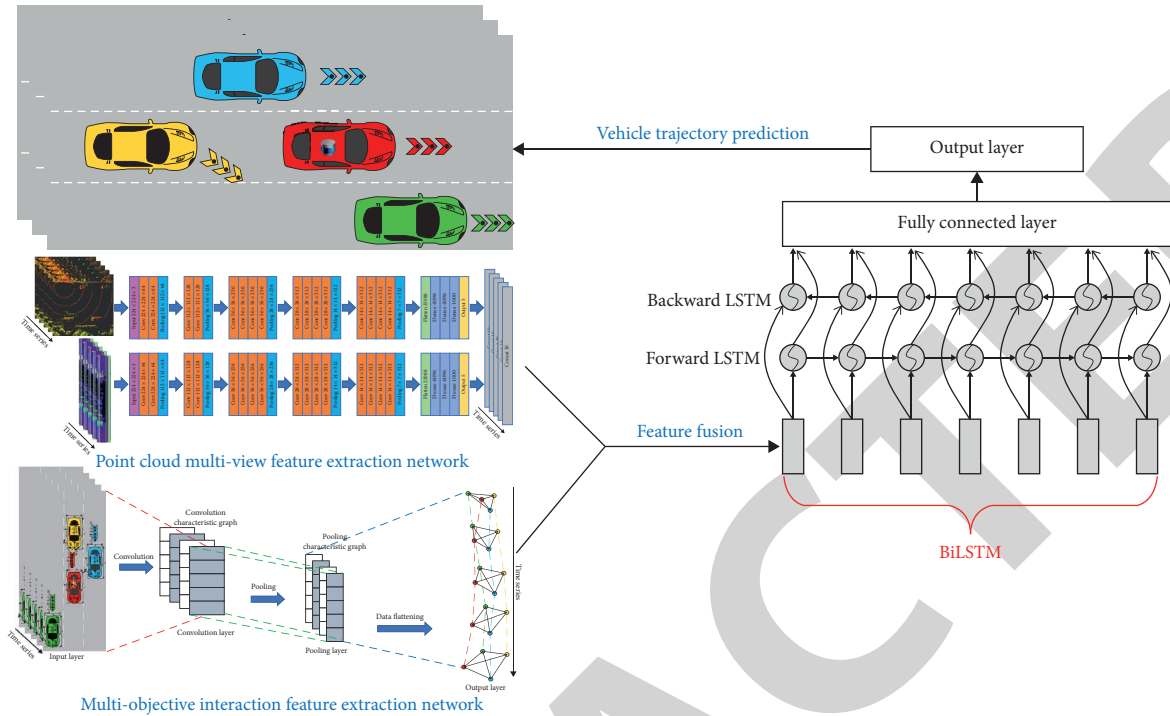
FIGURE 6: The overall network architecture of vehicle motion state prediction model integrating multiview features of point cloud and multitarget interactive information.
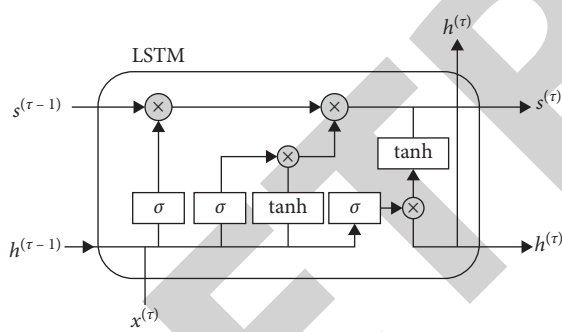

FIGURE 7: The LSTM structure diagram.

20 Hz. The 32-wire LiDAR can scan the surrounding environment with a vertical field of view angle of $-25°\sim15°$ and a horizontal angle of view of $360°$, with a maximum ranging range of 200 m and an output of 60×10^4 points per second, with the scanning frequency set to 20 Hz. The laptop is equipped with an Ubuntu 16.04 operating system, CUDA 9.0 deep learning parallel computing acceleration kit, independent graphics card NVIDIA GeForce GTX 1650 GPU, and Intel Core i5-9300H CPU with 2.4 GHz and 16 GB memory. The algorithm processor has built-in efficient environment detection-related algorithms. The Gigabit Ethernet switch ensures high-velocity data transmission of the data acquisition platform, and the uninterrupted power source provides a reliable power supply for the experimental acquisition equipment. The environmental point cloud data collected by LiDAR are sent to the

Gigabit Ethernet switch through an Ethernet cable and transmitted to the algorithm processor for environmental information detection. The results are sent to the notebook computer for storage and secondary operation processing visualization through Ethernet.

As shown in Figure 10, the test route is a two-way four-lane urban road from the Liuhe intersection of the East Second Ring Road to Nanzhou Bridge in Qixing District, Guilin, Guangxi. The route has a total length of 4.2 km, including a straight road and a curve with the approximate lengths of 3.6 and 0.6 km, respectively, and a velocity limit of 60 km/h. Before the test, the tester checked to ensure that the test vehicle and instruments were in good working condition. During the test, the tester drove from the starting point at Liuhe intersection to the endpoint (destination) of Nanzhou Bridge, which took approximately 7 minutes. In order to fully collect the point cloud scene data and vehicle interaction information of the road section, the tester drove the intelligent vehicle experimental platform back and forth 40 times to collect data from different target vehicles; to minimize the interaction between the front target vehicle and its surrounding vehicles during the data collection process of the intelligent vehicle experimental platform, the experimental data selected the data collected when the intelligent vehicle was driving behind the target vehicle and its surrounding vehicles and focused on extracting the scene data of single-vehicle operation, and interactions between two, three, and four vehicles for analysis. The visualization effect of the obtained data in the Robot Operating System (ROS) environment is shown in Figure 11.
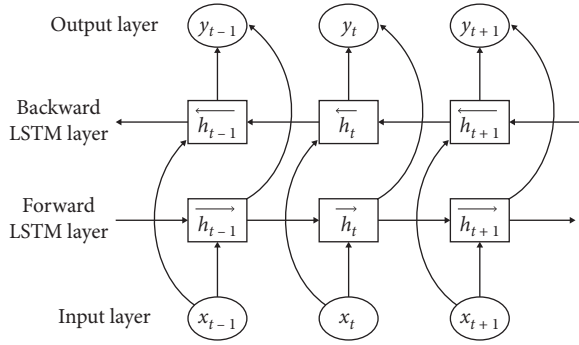
FIGURE 8: BiLSTM network structure.

The motion parameter information of surrounding traffic participants in complex scenes obtained by the intelligent vehicle experimental platform is shown in Table 1.

The intelligent vehicle experimental platform was returned 40 times, and the following data collection scenes of different target vehicles were screened and divided. The sample division of the training set and test set is shown in Table 2. In the different types of car-following scenarios, 100 groups of car-following data are used as the training set, and 30 groups of car-following data are used as the test set.

In order to verify the effectiveness of the proposed BiLSTM model, the model was implemented on the Keras deep learning platform based on TensorFlow and compared with FC, NN, and LSTM on the same dataset. The layer structure, output shape, and parameter number of the comparative experimental network model are shown in Table 3.

The difference between the predicted value and the real value must be calculated to effectively evaluate the effect of the target vehicle state prediction model. Generally, the mean square error (MSE) is used as the evaluation index for prediction results, and the calculation formula is shown as follows:

$$MSE = \frac{1}{n} \times \sum_{t=1}^{n} (A_t - F_t)^2, \qquad (10)$$

where A_t is the predicted value and F_t is the true value. The smaller the index MSE, the closer the predicted value is to the real value, which proves that the model has better performance and strong feature expression ability [45]. In order to effectively analyze the interaction and time-space relationship between vehicles in complex traffic scenes, this paper focuses on the prediction effect of vehicle global position X, global position Y, relative velocity V, and other parameters.

As shown in Figures 12(a)–12(d), the intelligent vehicle experimental platform follows the vehicle, and the targets are single-vehicle operations and two-, three-, and four-vehicle interactions. The MSE loss value of the global position X of the target vehicle is iterated 300 times. Results in the figure indicate that the effect of using the BiLSTM model is better than that of FC, NN, and LSTM models.

Figures 13(a)–13(d) show the intelligent vehicle experimental platform results using the MSE loss value of the global position X of the target vehicle iterated 300 times based on the point cloud multiview. The following scenario targets are single-vehicle operations and two-, three-, and four-vehicle interactions. The effect of using the BiLSTM model is better than that of FC, NN, and LSTM models. The statistical results are shown in Table 4. The MSE loss value of the global position X of the target vehicle in single-vehicle operations and two-, three-, and four-vehicle interaction scenarios based on point cloud multiview is lower than that in the single-vehicle operation and two-, three-, and four-vehicle interaction scenarios.

As shown in Figures 14(a)–14(d), the intelligent vehicle experimental platform follows the vehicle, and the targets are single-vehicle operations and two-, three-, and four-vehicle interactions. The MSE loss value of the global position Y of the target vehicle is iterated 300 times. As illustrated, the effect using the BiLSTM model is better than that of FC, NN, and LSTM models.

Figures 15(a)–15(d) show the following scene of the intelligent vehicle experimental platform where the MSE loss value of the global position Y of the target vehicle is iterated 300 times based on the multiview of the point cloud. The targets are single-vehicle operations and two-, three-, and four-vehicle interactions. The effect of using the BiLSTM model can be observed as better than that of the FC, NN, and LSTM models. The statistical results are shown in Table 5. The MSE loss value of the global position Y of the target vehicle in the single-vehicle operations and two-, three-, and four-vehicle interaction scenarios based on point cloud multiview is lower than that in the single-vehicle operations and two-, three-, and four-vehicle interaction scenarios.

As shown in Figures 16(a)–16(d), the intelligent vehicle experimental platform follows the vehicle, and the targets are single-vehicle operations and two-, three-, and four-vehicle interactions. The MSE loss value of the target vehicle relative velocity V is iterated 300 times. The effect of using the BiLSTM model is illustrated as better than that of FC, NN, and LSTM models.

Figures 17(a)–17(d) show the following scene of the intelligent vehicle experimental platform, where the MSE loss value of the target vehicle relative velocity V is iterated 300 times based on the multipoint cloud view. The target is single-vehicle operations and two-, three-, and four-vehicle interactions. As illustrated in the figure, the effect of using the BiLSTM model is better than FC, NN, and LSTM models. The statistical results are shown in Table 6. The MSE loss value of the target vehicle relative velocity V under the single-vehicle operations and two-, three-, and four-vehicle interaction scenarios based on point cloud multiview is lower than that of single-vehicle operations and two-, three-, and four-vehicle interaction scenarios.

As shown in Table 7, when the number of epoch is 300, the global position X average MSE, global position Y average MSE, and relative velocity V average MSE of BiLSTM model are significantly lower than those of other models; The average running time of FC model is the smallest, and the average running time of LSTM model is close to that of BiLSTM model. Considering the calculation cost of the model, it can be seen that BiLSTM model has the best timeliness.

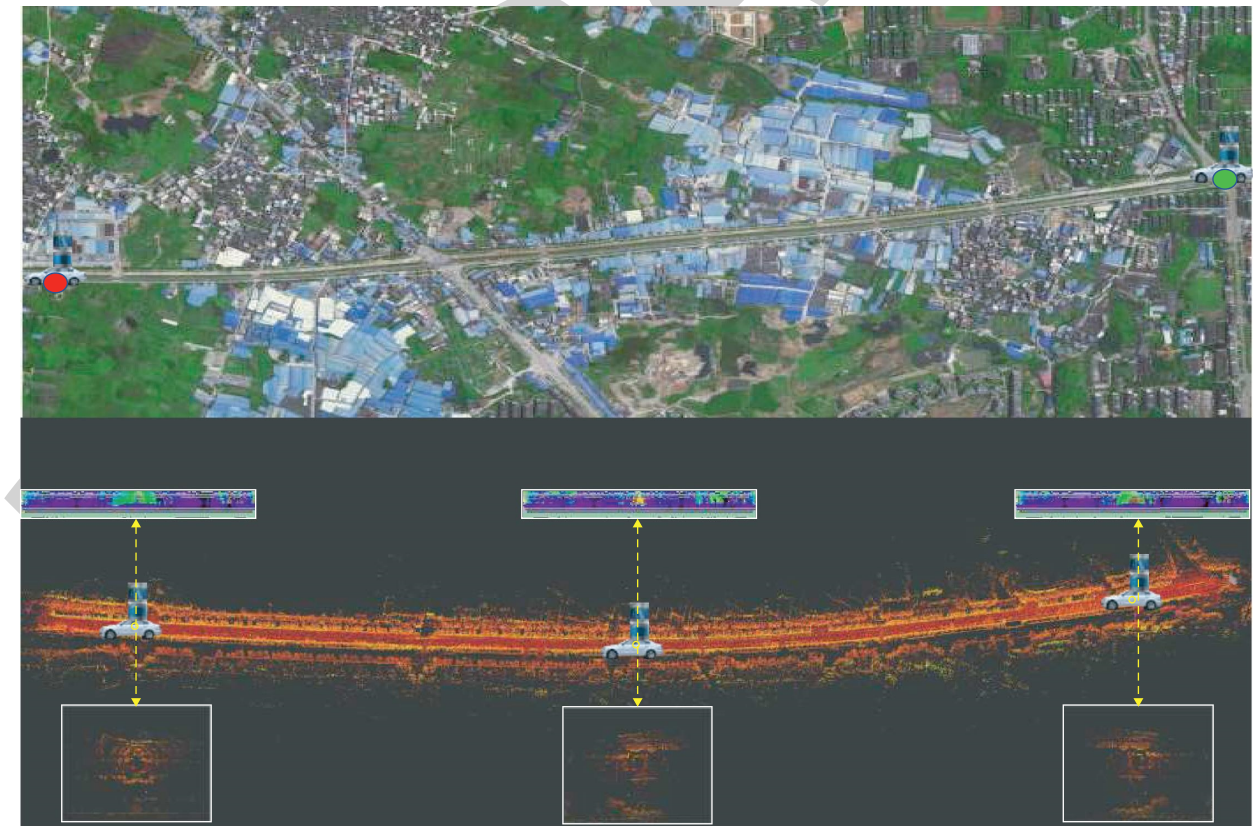Figure 9: Intelligent vehicle test platform.

Figure 10: Test route and road information collection.
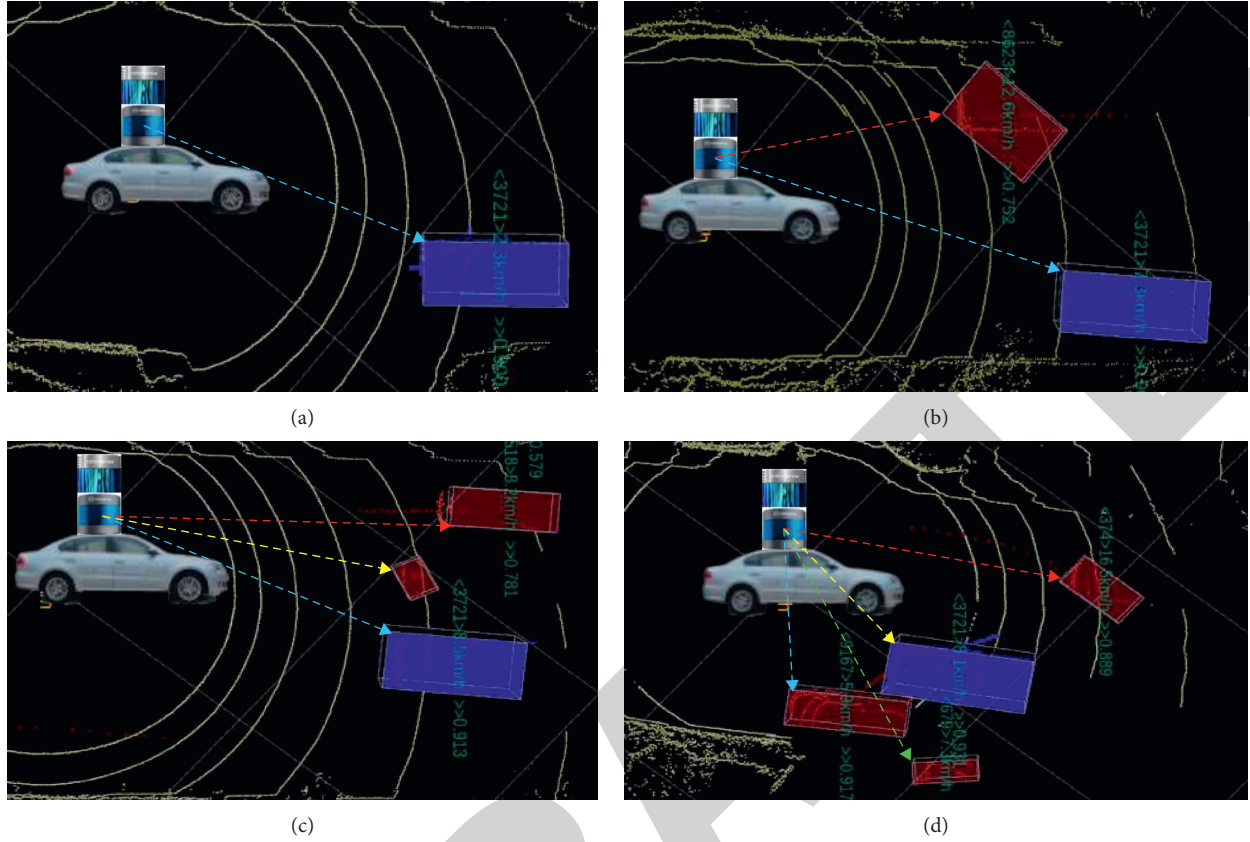
(a)

(b)

(c)

(d)

Figure 11: Vehicle interaction scenario. (a) Single-vehicle operation. (b) Two-vehicle interaction. (c) Three-vehicle interaction. (d) Four-vehicle interaction.

Table 1: Motion parameter information of surrounding traffic participants in complex scenes.

Information category	Describe	Unit
Time	Vehicle LiDAR data acquisition time	s
Pointcloud	Current frame environment point cloud	—
Timestamp	Timestamp when the target is perceived	s
ID	Tracker ID when the perceived target is tracked	
Location	Position of the target center in the global coordinate system	m
Length × width × height	Target bounding box size	m
Nearest point	Coordinates of the point closest to the radar in the target bounding box in the global coordinate system	—
Velocity	Velocity vector of target in global coordinate system	m/s

Table 2: Data set division.

Car following scene	Training set	Test set
Single-vehicle operation	100	30
Two-vehicle interaction	100	30
Three-vehicle interaction	100	30
Four-vehicle interaction	100	30

In order to further study the relationship between the prediction algorithm model, point cloud multiview, and multivehicle interaction information, the prediction effects of FC, NN, LSTM, and BiLSTM models on the same target vehicle state in the same scene were analyzed. As shown in Figures 18(a)–18(c), the dotted line of the predicted value of the red BiLSTM model is best matched with the solid line of

the blue real value. Therefore, the BiLSTM model is superior to FC, NN, and LSTM models in predicting the state of target vehicles.

The prediction effect of the BiLSTM model on the target vehicle state was further analyzed when different numbers of vehicles interact in the same scene. As shown in Figures 19(a)–19(c), the dotted line of the predicted value of the green

TABLE 3: Comparative experimental network model.

Network model type	Layer structure	Output shape	Number of parameters
FC	Flatten	None, 160	0
	Dense	None, 8	1288
	Dense	None, 1	9
NN	Dens	None, 32	192
	Dense	None, 32	1056
	Dense	None, 32	1056
	Dense	None, 1	33
LSTM	LSTM	None, 128	66,560
	Dense	None, 1	129
BiLSTM	Bidirectional	None, 256	133,120
	Dense	None, 1	257
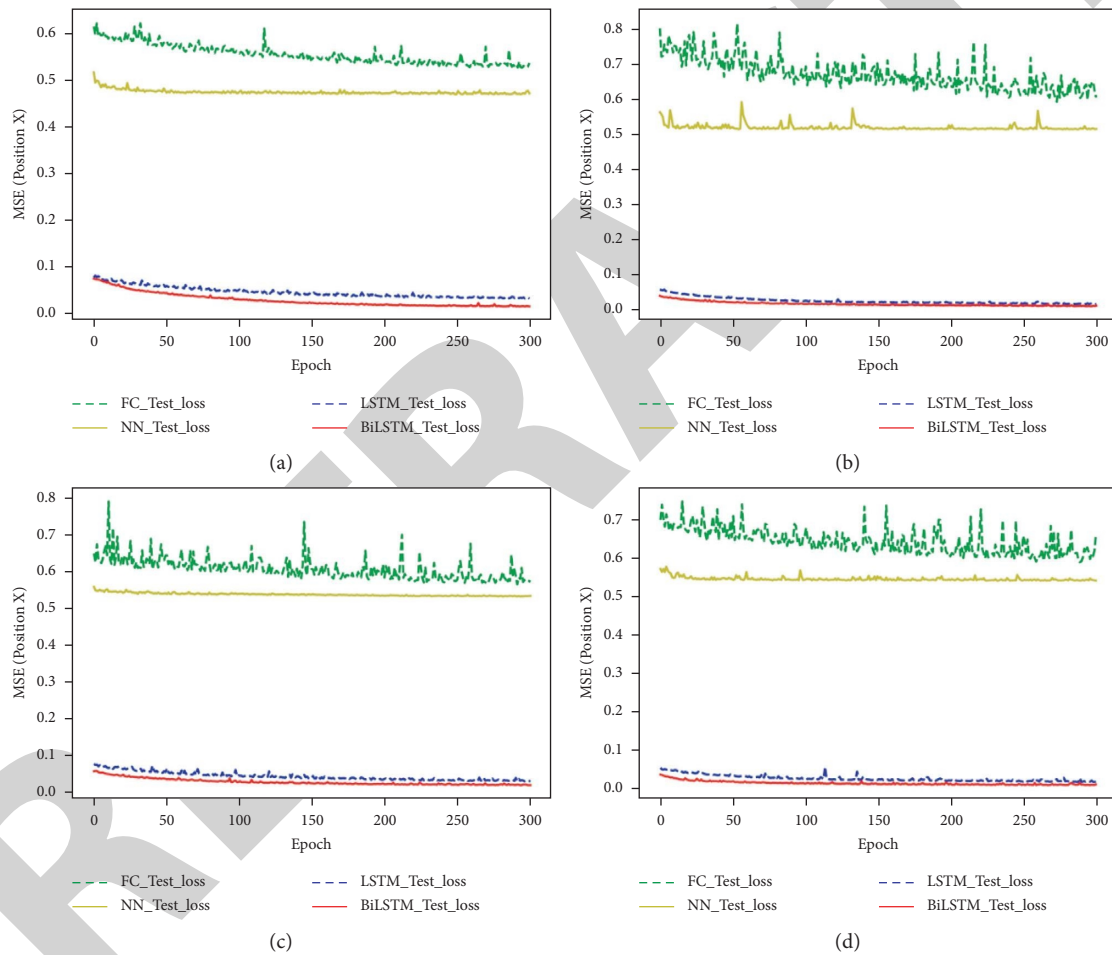
FIGURE 12: MSE loss value of the target vehicle global position X in different interaction scenarios. (a) Single-vehicle operation. (b) Two-vehicle interaction. (c) Three-vehicle interaction. (d) Four-vehicle interaction.

single-vehicle interaction scene is best fit with the solid line of the blue real value. Therefore, the prediction effect of the BiLSTM model on the target vehicle state in the single-vehicle interaction scene is better than that of multivehicle interaction.

The prediction effect of the BiLSTM model based on point cloud multiview on the target vehicle state when different numbers of vehicles interact in the same scene was

also analyzed. As shown in Figures 20(a)–20(c), the green dotted line of the predicted value based on the point cloud multiview single-vehicle interaction scene fits best with the solid blue line of the real value. Therefore, the prediction effect of the BiLSTM model on the target vehicle state of single-vehicle interaction scene based on point cloud multiview is better than that of multivehicle transportation.
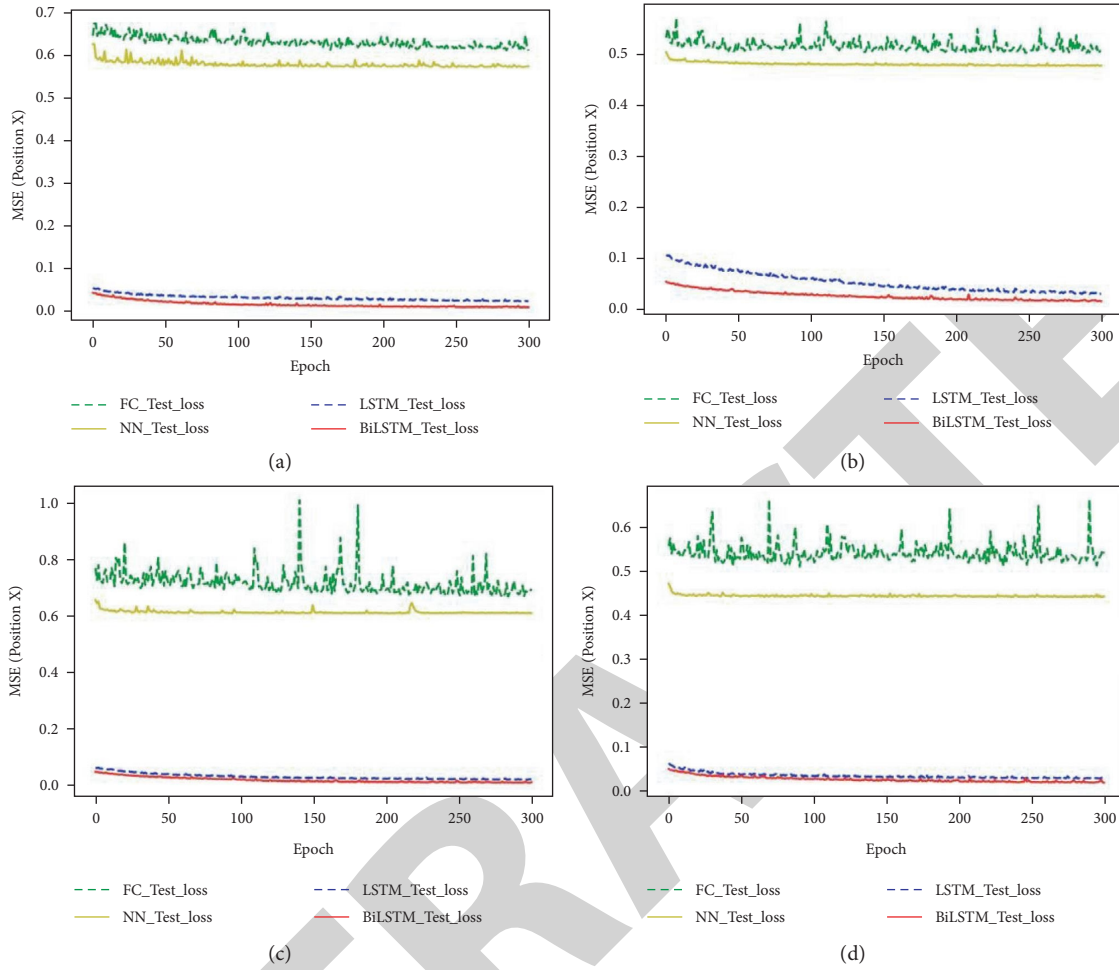
Figure 13: MSE loss value of the global position X of the target vehicle based on point cloud multiview in different interaction scenarios. (a) Single-vehicle operation + multiview. (b) Two-vehicle interaction + multiview. (c) Three-vehicle interaction + multiview. (d) Four-vehicle interaction + multiview.

Table 4: MSE statistics of target vehicle global position X of each prediction model in different interaction scenarios and based on point cloud multiview.

Combination type of scenario and multiview	FC	NN	LSTM	BiLSTM
Single-vehicle operation	0.5567	0.4628	0.0184	0.0172
Single-vehicle operation + multiview	0.4764	0.4001	0.0145	0.0115
Two-vehicle interaction	0.7157	0.5625	0.0363	0.0248
Two-vehicle interaction + multiview	0.6997	0.5014	0.0139	0.0097
Three-vehicle interaction	0.6205	0.5619	0.0282	0.0225
Three-vehicle interaction + multiview	0.4870	0.4664	0.0157	0.0099
Four-vehicle interaction	0.5988	0.4604	0.0470	0.0358
Four-vehicle interaction + multiview	0.5102	0.4392	0.0180	0.0096

It can be seen from the comparison between Figures 19 and 20 that the fitting degree of the dotted line of the predicted value based on point cloud multiview and vehicle interaction information of each color with the solid line of the blue real value is higher than that of the dotted line of the predicted value based on vehicle interaction information with the solid line of the blue real value only. Therefore, the effect of the BiLSTM model on target vehicle state prediction based on point cloud multiview and vehicle interaction information is better than that based on vehicle interaction information only.
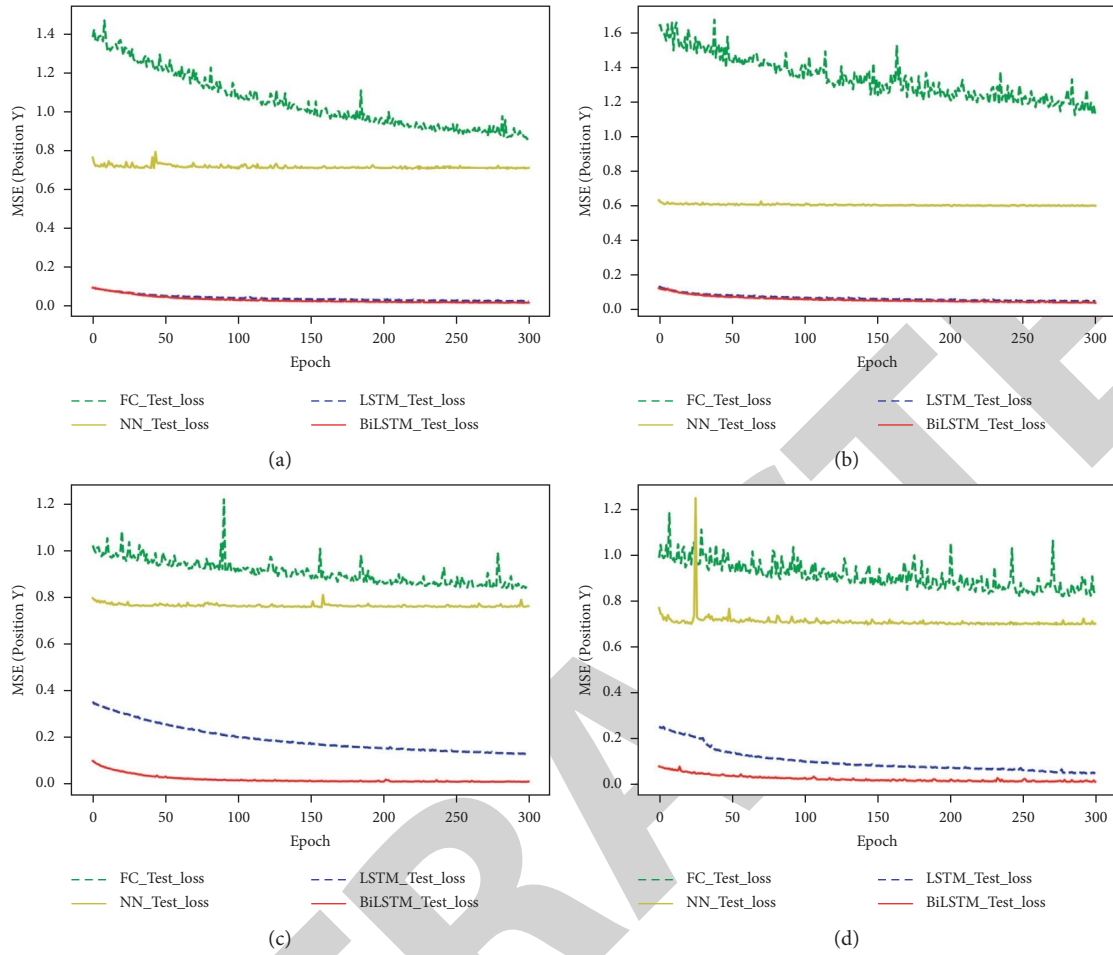
(a)

(b)

(c)

(d)

Figure 14: MSE loss value of global position *Y* of the target vehicle in different interaction scenarios. (a) Single-vehicle operation. (b) Two-vehicle interaction. (c) Three-vehicle interaction. (d) Four-vehicle interaction.
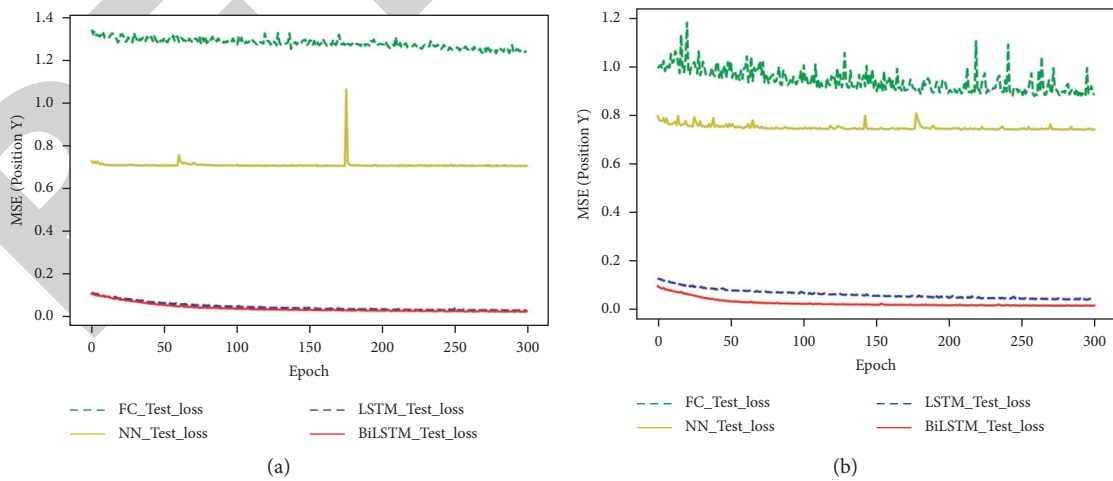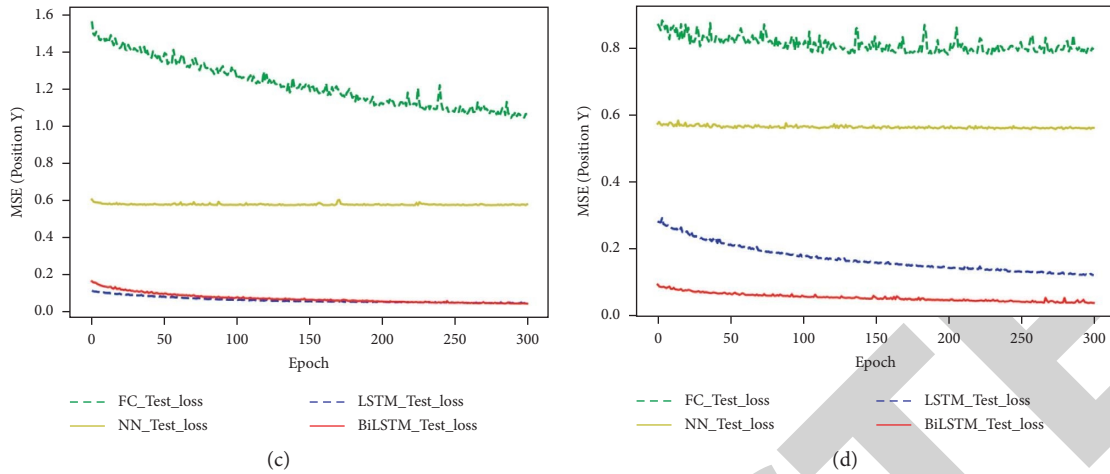

(a)

(b)

Figure 15: Continued.

(c)

(d)

FIGURE 15: MSE loss value of global position Y of the target vehicle based on point cloud multiview in different interaction scenarios. (a) Single-vehicle operation + multiview. (b) Two-vehicle interaction + multiview. (c) Three-vehicle interaction + multiview. (d) Four-vehicle interaction + multiview.

TABLE 5: MSE statistics of the target vehicle global position Y of each prediction model in different interaction scenarios and based on point cloud multiview.

Combination type of scenario and multiview	FC	NN	LSTM	BiLSTM
Single-vehicle operation	0.8603	0.8059	0.0442	0.0324
Single-vehicle operation + multiview	0.8206	0.8041	0.0311	0.0262
Two-vehicle interaction	0.9565	0.8628	0.0753	0.0584
Two-vehicle interaction + multiview	0.7057	0.6706	0.0391	0.0301
Three-vehicle interaction	1.1467	0.9292	0.0674	0.0482
Three-vehicle interaction + multiview	0.8440	0.7958	0.0282	0.0134
Four-vehicle interaction	1.0361	0.8102	0.0729	0.0648
Four-vehicle interaction + multiview	0.9250	0.6871	0.0606	0.0182

(a)

(b)
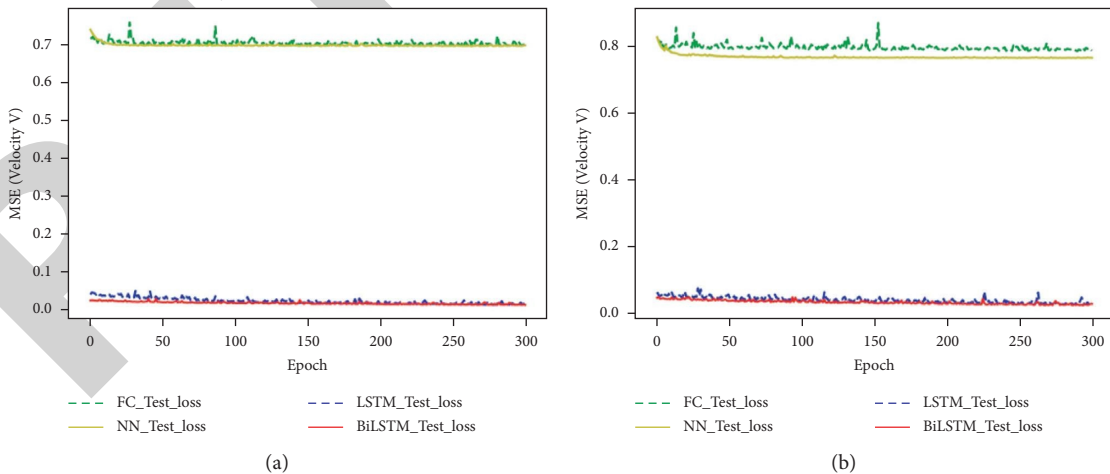
FIGURE 16: Continued.

(c)

(d)
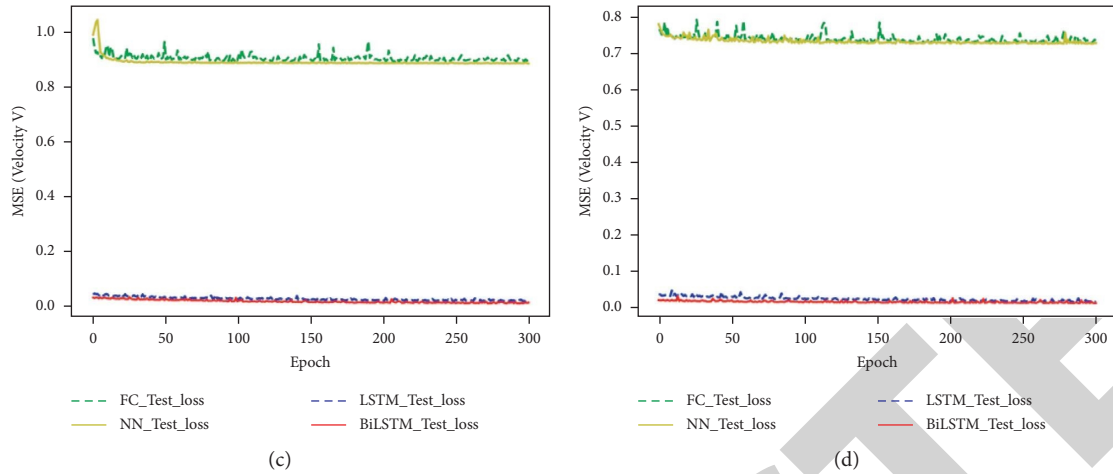
Figure 16: MSE loss value of the target vehicle relative velocity V in different interaction scenarios. (a) Single-vehicle operation. (b) Two-vehicle interaction. (c) Three-vehicle interaction. (d) Four-vehicle interaction.
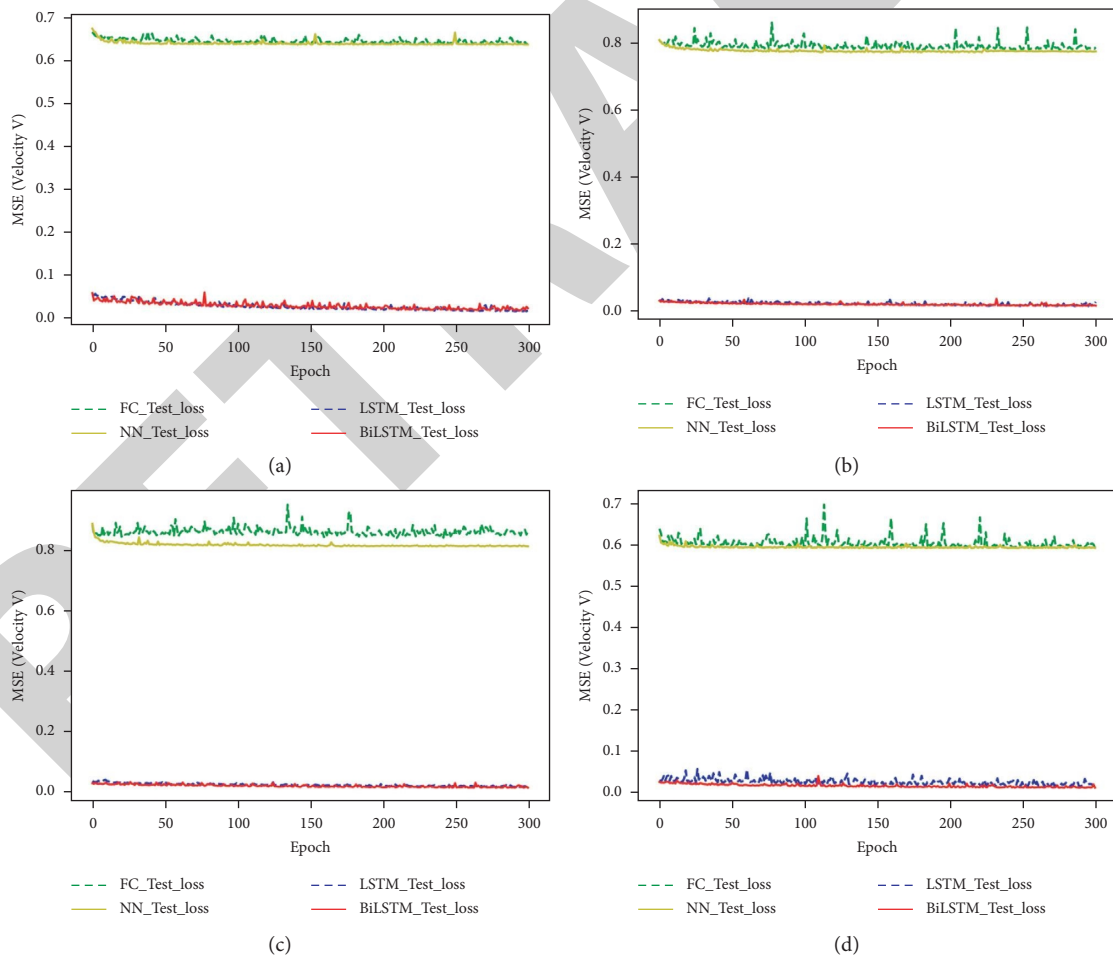

(a)

(b)

(c)

(d)

Figure 17: MSE loss value of the target vehicle relative velocity V based on point cloud multiview in different interaction scenarios. (a) Single-vehicle operation + multiview. (b) Two-vehicle interaction + multiview. (c) Three-vehicle interaction + multiview. (d) Four-vehicle interaction + multiview.

TABLE 6: MSE statistics of the target vehicle relative velocity V of each prediction model in different interaction scenarios and based on point cloud multiview.

Combination type of scenario and multiview	FC	NN	LSTM	BiLSTM
Single-vehicle operation	0.7510	0.7291	0.0115	0.0110
Single-vehicle operation + multiview	0.6296	0.6175	0.0108	0.0103
Two-vehicle interaction	0.7441	0.7281	0.0244	0.0174
Two-vehicle interaction + multiview	0.7385	0.7236	0.0164	0.0161
Three-vehicle interaction	0.7653	0.7473	0.0244	0.0156
Three-vehicle interaction + multiview	0.6538	0.6496	0.0148	0.0121
Four-vehicle interaction	0.7193	0.7166	0.0238	0.0210
Four-vehicle interaction + multiview	0.6880	0.6543	0.0179	0.0155

TABLE 7: Performance comparison statistics of prediction model.

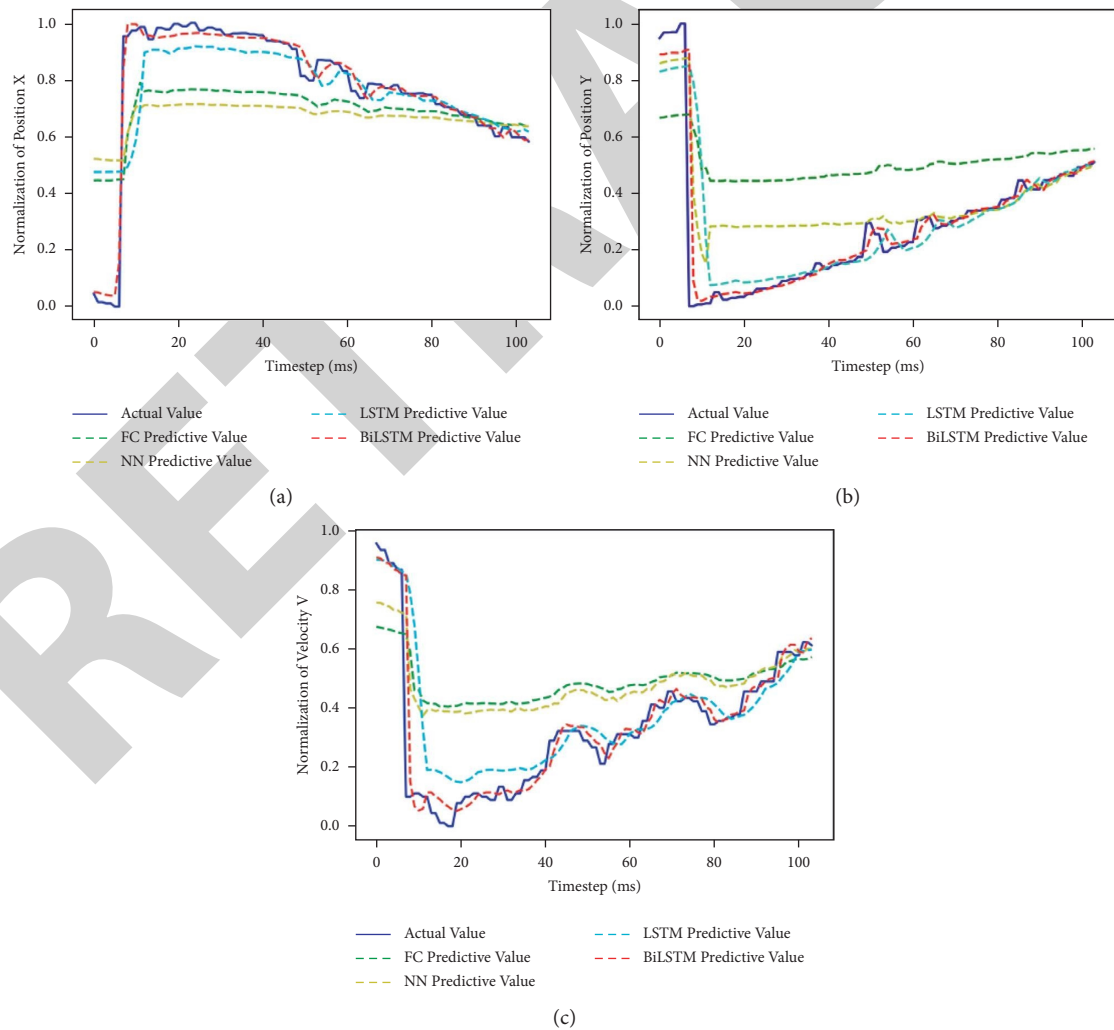| Model parameter	FC	NN	LSTM	BiLSTM
Global position X average MSE	0.5831	0.4818	0.0240	0.0176
Global position Y average MSE	0.9118	0.7957	0.0520	0.0364
Relative velocity V average MSE	0.7112	0.6957	0.0180	0.0148
Average running time (s)	0.0318	0.0475	0.0726	0.0774

(a)

(b)

(c)

FIGURE 18: Prediction effect of the model on the state of the same target vehicle in the same scene. (a) Position X predictive value. (b) Position Y predictive value. (c) Velocity V predictive value.
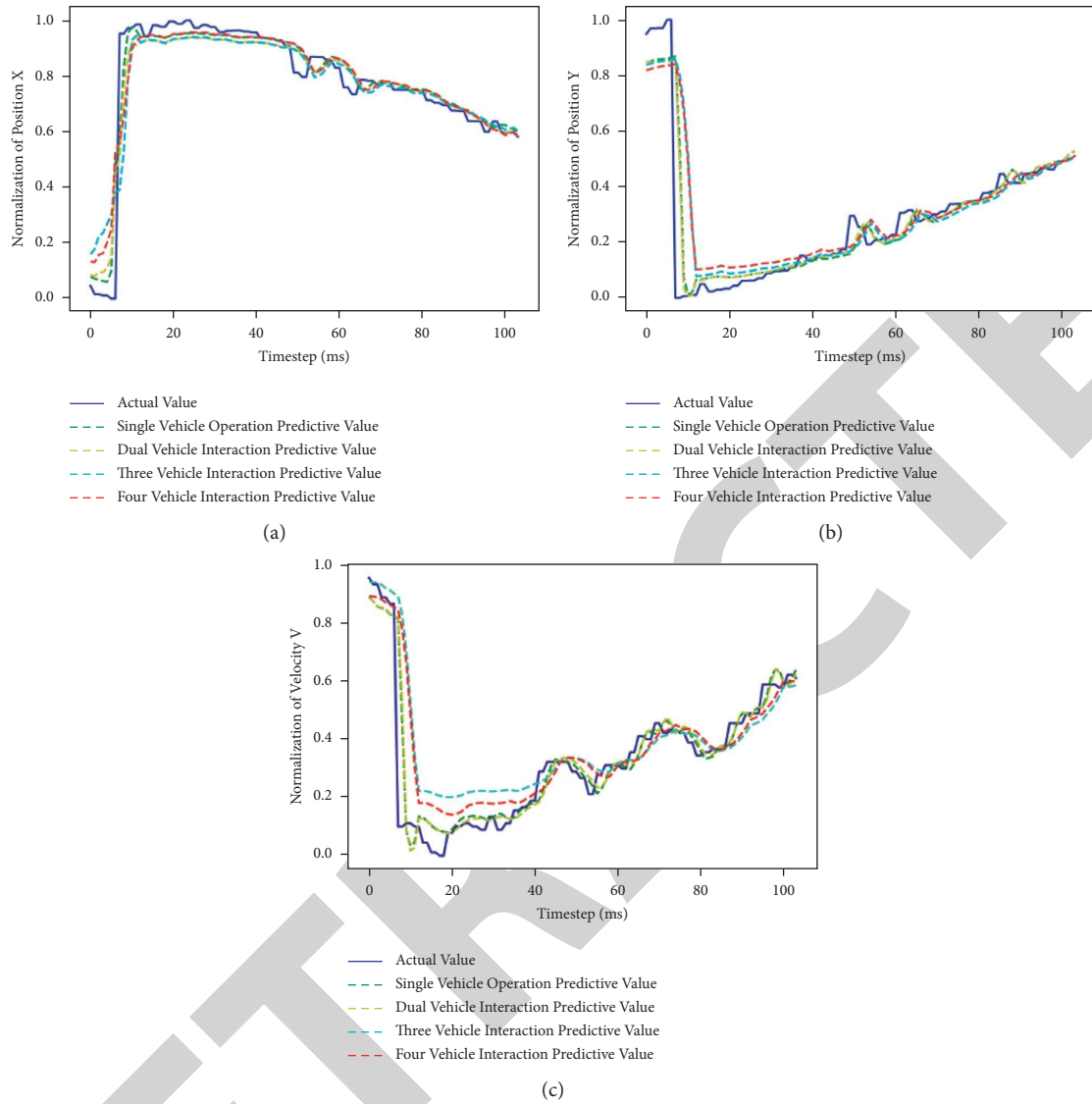
(a)

(b)

(c)

Figure 19: Prediction effect of the BiLSTM model on the target vehicle state when different numbers of vehicles interact in the same scene. (a) Position X predictive value. (b) Position Y predictive value. (c) Velocity V predictive value.
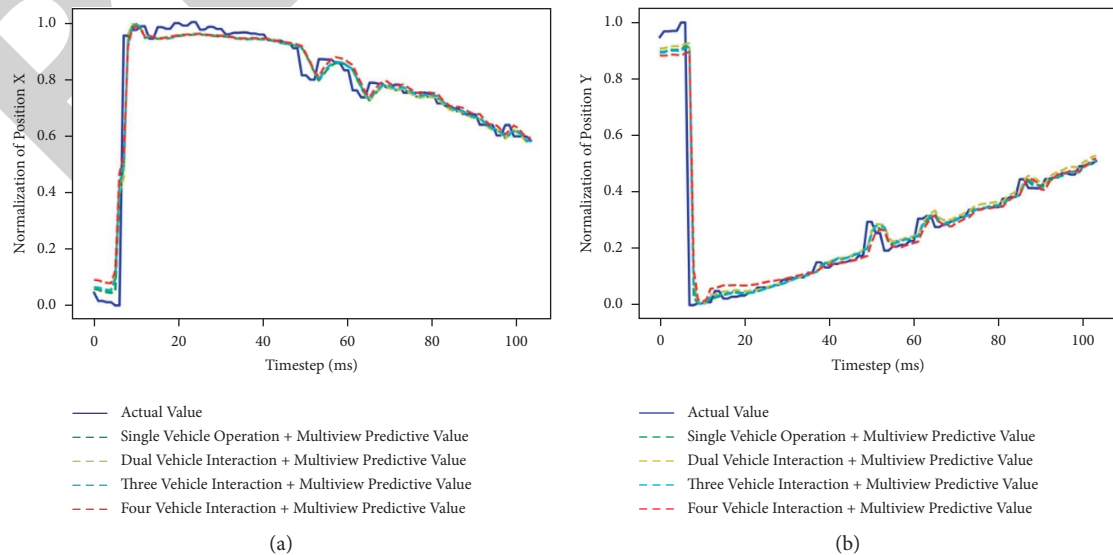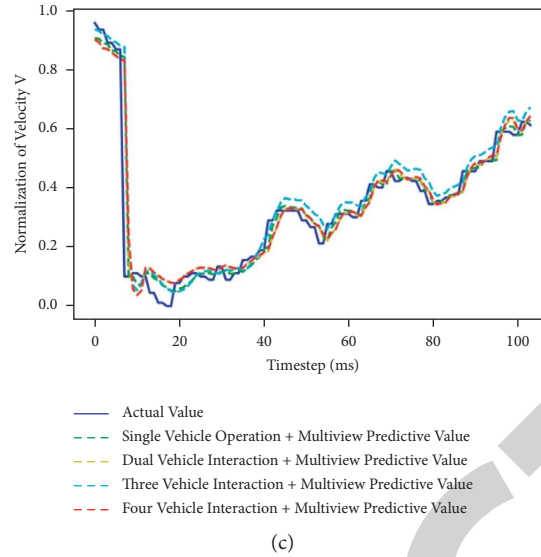

(a)

(b)

Figure 20: Continued.

(c)

FIGURE 20: Prediction effect of the BiLSTM model on the target vehicle state in different interaction scene based on point cloud multiview. (a) Position *X* predictive value. (b) Position *Y* predictive value. (c) Velocity *V* predictive value.

6. Conclusions

A complex traffic environment perception technology based on two multiline LiDAR was proposed to effectively predict the motion state of traffic participants around intelligent vehicles in complex scenes. A vehicle motion state prediction algorithm integrating point cloud timing multiview features and multitarget interactive information was proposed to analyze the influence of the target vehicle motion state affected by the interaction between the surrounding environment and objects. With the help of real-time point cloud information perceived by the LiDAR, the time sequence aerial view map and the time sequence front view depth map were obtained, and the time sequence high-level abstract combination features in the multiview scene were extracted using an improved VGG19 network model. Both were fused with the potential spatiotemporal interaction features of extracting the multitarget operation state data detected by the laser radar using a one-dimensional convolution neural network. The temporal feature vector was constructed as the input data of the BiLSTM network, and the desired input-output mapping relationship was trained to predict the motion state of traffic participants. The test results showed that the prediction effect of state parameters, such as global position *X*, global position *Y*, and relative velocity *V* of the target vehicle using the BiLSTM model, was better than the FC, NN, and LSTM models. The prediction effect of the target vehicle state based on point cloud multiview and vehicle interaction information was also better than that based on vehicle interaction information only. How to effectively evaluate the risk of intelligent vehicle operation environment remains a challenge. In future work, we plan to research the prediction of the operational risk field of intelligent vehicles based on dual multiline LiDAR.

Data Availability

The data used to support the findings of this study have not been made available due to data privacy.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this study.

References

[1] Y. Liang, S. Muller, D. Schwendner, D. Rolle, D. Ganesch, and I. Schaffer, "A scalable framework for robust vehicle state estimation with a fusion of a low-cost IMU, the GNSS, radar, a camera and lidar," in *Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, Portland, OR, USA, April 2020.

[2] U. H. Kim, D. Ka, H. Yeo, and J.-H. Kim, *A Real-Time Vision Framework for Pedestrian Behavior Recognition and Intention*

Prediction at Intersections Using 3D Pose Estimation, IEEE, Daejeon, South Korea, 2020.

[3] N. Lyu, J. Wen, Z. Duan, and C. Wu, "Vehicle trajectory prediction and cut-in collision warning model in a connected vehicle environment," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–16, 2020.

[4] J. Huang, S. Xiao, L. Xiao, Y. Jiao, Z. Qiang Wang, and X. Wang, "Dynamic model construction and simulation based on intelligent driving vehicle," in *Proceedings of the 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, IEEE, Chongqing, China, June 2021.

[5] M. N. Azadani and A. Boukerche, "Driving behavior analysis guidelines for intelligent transportation systems," *IEEE Transactions on Intelligent Transportation Systems*, no. 99, pp. 1–19, 2021.

[6] Q. Jiang et al., "Intention of risk-taking behavior at unsignalized intersections under the connected vehicle environment," *IEEE Access*, no. 99, p. 1, 2021.

[7] S. Ammoun and F. Nashashibi, "Real time trajectory prediction for collision risk estimation between vehicles," in *Proceedings of the IEEE 5th International Conference on. 2009*, IEEE, Cluj-Napoca, Romania, August 2009.

[8] J. Wiest et al., "Probabilistic trajectory prediction with Gaussian mixture models," in *Proceedings of the Intelligent Vehicles Symposium (IV)*, June 2012.

[9] L. L. Zhang, H. Zhu, C.-Z. Wu, and A.-W. Zheng, "Risk estimation to vehicles collision at intersection based on ARMA prediction model," *Journal of Transportation Systems Engineering and Information Technology*, 2015.

[10] B. Kim and K. Yi, "Probabilistic and holistic prediction of vehicle states using sensor fusion for application to integrated vehicle safety systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2178–2190, 2014.

[11] B. Kim and K. Yi, "Probabilistic states prediction algorithm using multi-sensor fusion and application to Smart Cruise Control systems," in *Proceedings of the Intelligent Vehicles Symposium (IV)*, IEEE, Gold Coast, QLD, Australia, June 2013.

[12] M. Schreier, V. Willert, and J. Adamy, "An integrated approach to maneuver-based trajectory prediction and criticality assessment in arbitrary road environments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 10, pp. 2751–2766, 2016.

[13] D. N. Jagadish, A. Chauhan, and L. Mahto, "Autonomous vehicle path prediction using conditional variational autoencoder networks," *PAKDD*, vol. 12712, 2021.

[14] M. Neumeier, M. Betsch, A. Tollkuhn, and T. Berberich, "Variational autoencoder-based vehicle trajectory prediction with an interpretable latent space," *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, IEEE, in *Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, September 2021.

[15] J. Schreier and A. Matthias, "Bayesian environment representation, prediction, and criticality assessment for driver assistance systems," *At - Automatisierungstechnik*, vol. 65, no. 2, 2017.

[16] M. Biparva, David Fernández-Llorca, Rubén Izquierdo-Gonzalo, and John K. Tsotsos, *Video Action Recognition for Lane-Change Classification and Prediction of Surrounding Vehicles*, IEEE, Llorca, 2021.

[17] D. Fernández-Llorca, Mahdi Biparva, Ruben Izquierdo-Gonzalo, and John K. Tsotsos, "Two-stream networks for lane-change prediction of surrounding vehicles," in *Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Rhodes, Greece, September 2020.

[18] T. Quan and J. Firl, "Modelling of traffic situations at urban intersections with probabilistic non-parametric regression," in *Proceedings of the Intelligent Vehicles Symposium*, IEEE, Gold Coast, QLD, Australia, June 2013.

[19] A. Patterson, A. Gahlawat, and N. Hovakimyan, *Learning Probabilistic Intersection Traffic Models for Trajectory Prediction*, Urbana, 2020.

[20] I. V. Pustokhina, D. A. Pustokhin, T. Vaiyapuri, D. Gupta, S. Kumar, and K. Shankar, "An automated deep learning based anomaly detection in pedestrian walkways for vulnerable road users safety," *Safety Science*, vol. 142, p. 105356, 2021.

[21] D. Muley, M. Ghanim, and M. Kharbeche, "Prediction of traffic conflicts at signalized intersections using SSAM," *Procedia Computer Science*, vol. 130, pp. 255–262, 2018.

[22] T. Streubel and K. H. Hoffmann, "Prediction of driver intended path at intersections," in *Proceedings of the 2014 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, Dearborn, MI, USA, June 2014.

[23] T. Gindele, S. Brechtel, and R. Dillmann, "Learning driver behavior models from traffic observations for decision making and planning," *IEEE Intelligent Transportation Systems Magazine*, vol. 7, no. 1, pp. 69–79, 2015.

[24] K. Shu, H. Yu, X. Chen et al., "Autonomous driving at intersections: a behavior-oriented critical-turning-point Approach for decision making," *IEEE*, no. 99, , 2021.

[25] B. D. Kim et al., "Probabilistic vehicle trajectory prediction over occupancy grid map via recurrent neural network," in *Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, Yokohama, Japan, October 2017.

[26] N. Deo and M. M. Trivedi, "Multi-modal trajectory prediction of surrounding vehicles with maneuver based LSTMs," in *Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, Changshu, China, June 2018.

[27] M. M. Aliabadi, H. Emami, M. Dong, and Y. Huang, "Attention-based recurrent neural network for multistep-ahead prediction of process performance," *Computers & Chemical Engineering*, vol. 140, p. 106931, 2020.

[28] A. Ip, L. Irio, and R. Oliveira, "Vehicle trajectory prediction based on LSTM recurrent neural networks," in *Proceedings of the 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)*, IEEE, Helsinki, Finland, April 2021.

[29] S. Mukherjee, S. Wang, and A. Wallace, "Interacting vehicle trajectory prediction with convolutional recurrent neural networks," in *Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Paris, France, May 2020.

[30] C. Tang, J. Chen, and M. Tomizuka, "Adaptive probabilistic vehicle trajectory prediction through physically feasible bayesian recurrent neural network," in *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, IEEE, Montreal, QC, Canada, May 2019.

[31] S. Eiffert, K. Li, M. Shan, S. Worrall, S. Sukkarieh, and E. Nebot, "Probabilistic crowd gan: multimodal pedestrian trajectory prediction using a graph vehicle-pedestrian attention network," *IEEE Robotics and Automation Letters*, vol. 5, no. 4, pp. 5026–5033, 2020.

[32] M. Lisotto, P. Coscia, and L. Ballan, "Social and scene-aware trajectory prediction in crowded spaces," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*

Workshop (ICCVW), IEEE, Seoul, Korea (South), October 2019.

[33] X. Shi et al., "Multimodal interaction-aware trajectory prediction in crowded space," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 7, pp. 11982–11989, 2020.

[34] S. Zamboni et al., "Pedestrian trajectory prediction with convolutional neural networks," *Pattern Recognition*, vol. 121, p. 108252, 2021.

[35] H. Zhang et al., *A Multi-Modal States Based Vehicle Descriptor and Dilated Convolutional Social Pooling for Vehicle Trajectory Prediction*, SAE International, United States, 2020.

[36] A. Alahi et al., "Social LSTM: human trajectory prediction in crowded spaces," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, June 2016.

[37] A. Vemula, K. Muelling, and J. Oh, "Social attention: modeling attention in human crowds," in *Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, Brisbane, QLD, Australia, May 2018.

[38] A. Jain et al., "Structural-RNN: deep learning on spatio-temporal graphs," in *Proceedings of the IEEE Conference on Computer Vision & Pattern Recognition*, IEEE, Las Vegas, NV, USA, June 2016.

[39] A. Sadeghian et al., "CAR-net: clairvoyant attentive recurrent network," in *Proceedings of the Computer Vision – ECCV 2018,Lecture Notes in Computer Science*, pp. 162–180, Springer, Cham, Switzerland AG, October 2018.

[40] X. Hao, Q. H. Du, and M. Reynolds, "A location-velocity-temporal attention LSTM model for pedestrian trajectory prediction," *IEEE Access*, no. 99, p. 1, 2020.

[41] A. Sadeghian et al., "SoPhie: an attentive gan for predicting paths compliant to social and physical constraints," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Long Beach, CA, USA, June 2019.

[42] R. W. Liu, J. Nie, S. Garg, Z. Xiong, Y. Zhang, and M. S. Hossain, "Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5374–5385, 2021.

[43] M. Y. R. W. Liu, "MVFFNet: multi-view feature fusion network for imbalanced ship classification," *Pattern Recognition Letters*, vol. 151, pp. 26–32, 2021.

[44] T. Liang, Z. Meng, G. Xie, and S. Fan, "Multi-running state health assessment of wind turbines drive system based on BiLSTM and GMM," *IEEE Access*, vol. 8, pp. 143042–143054, 2020.

[45] L. Deng et al., "Approximation of bias and mean-squared error in two-sample Mendelian randomization analyses," *Biometrics*, vol. 76, no. 2, 2020.